Clinical Radiology xxx (2017) 1–7

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks

D.H. Kim*, T. MacKinnon

Medical Imaging Department, Royal Devon and Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK

ARTICLE INFORMATION

Article history:
Received 28 May 2017
Accepted 14 November 2017

AIM: To identify the extent to which transfer learning from deep convolutional neural networks (CNNs), pre-trained on non-medical images, can be used for automated fracture detection on plain radiographs.

MATERIALS AND METHODS: The top layer of the Inception v3 network was re-trained using lateral wrist radiographs to produce a model for the classification of new studies as either “fracture” or “no fracture”. The model was trained on a total of 11,112 images, after an eightfold data augmentation technique, from an initial set of 1,389 radiographs (695 “fracture” and 694 “no fracture”). The training data set was split 80:10:10 into training, validation, and test groups, respectively. An additional 100 wrist radiographs, comprising 50 “fracture” and 50 “no fracture” images, were used for final testing and statistical analysis.

RESULTS: The area under the receiver operator characteristic curve (AUC) for this test was 0.954. Setting the diagnostic cut-off at a threshold designed to maximise both sensitivity and specificity resulted in values of 0.9 and 0.88, respectively.

CONCLUSION: The AUC scores for this test were comparable to state-of-the-art providing proof of concept for transfer learning from CNNs in fracture detection on plain radiographs. This was achieved using only a moderate sample size. This technique is largely transferable, and therefore, has many potential applications in medical imaging, which may lead to significant improvements in workflow productivity and in clinical risk reduction.

Introduction

Demands on radiology services have increased dramatically in recent years causing a considerable strain on the workforce. In the UK, the number of computed tomography (CT) examinations increased by 29% between 2012 and 2015, whilst recruitment lagged behind, leaving 9% of consultant radiology posts vacant in 2015. Furthermore, in the same year nearly all radiology departments declared a failure to meet their reporting requirements.[1] In February 2016, there was an estimated backlog of 200,000 plain radiographs and 12,000 cross-sectional studies.[2] It is clear from these figures that improvements in reporting efficiency and workflow management are desperately needed if we are to avoid patient harm from delayed or missed diagnosis.

Artificial intelligence (AI) has the potential to address these issues. The wide adoption of electronic picture

* Guarantor and correspondent: D. H. Kim, Royal Devon and Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK. Tel.: 01392 411611; Mob.: 07799113665.

E-mail addresses: Daniel.kim@nhs.net, dan_kim92@hotmail.com (D.H. Kim).

https://doi.org/10.1016/j.crad.2017.11.015
0009-9260/

archiving and communications systems (PACS) has resulted in the development of one of the largest image data sets in existence. In the UK, there were 41 million imaging studies performed in 2016 alone.[3] These data lend themselves perfectly to machine learning.

Machine learning is a form of AI, which uses algorithms that iteratively improve, or learn, in response to training data in order to make autonomous predictions.[4] Supervised machine learning is one subtype, which relies on the provision of pre-labelled training data. In the field of general imaging and computer vision, deep learning is the leading machine learning tool.[5] Deep learning refers to techniques that build on developments in artificial neural networks in which multiple network layers are added to increase the levels of abstraction and performance.[6]

A simplified overview of the mechanics behind deep convolutional neural networks (CNNs) is provided in Fig 1. Neural networks aim to mimic the structure of the human brain. They use a series of interconnected "neurons", which collect, summarise, and/or transform data before passing the values to the next neuron in the sequence. This process culminates in an output layer, which can be used to formulate a prediction.

Constructing and training an effective neural network from scratch requires huge amounts of data. State-of–the-art image classification networks are frequently trained on data sets containing millions of images, facilitated by multiple computer servers, running continuously for several weeks.[9] This is not feasible for the majority of medical researchers. One method for overcoming this is to use a process called transfer learning. This is the process of adopting powerful and highly refined features from large existing pre-trained CNNs and using these as a starting point in training a new model for a different task. These low-level features, such as those illustrated in Fig 2, can be thought of as the basic building blocks of images such as lines and curves, and have been shown to be applicable to many different image-recognition tasks.[10] This technique can vastly reduce the computational requirements needed for network training and can deliver substantial performance benefits compared with training a CNN from scratch.[11]

Transfer learning has been relatively underutilised in the clinical setting despite the availability of vast amounts of image data. Two recent publications have demonstrated that transfer learning from pre-trained CNNs can produce human-level diagnostic results in the categorisation of skin lesions and in defining disease on digital retinal images.[12,13] Similar levels of diagnostic accuracy have yet to be reported in the analysis of plain radiographs; however, a few promising studies have recently been published. For example, one study retrained the GoogLeNet CNN for the detection of pathology in plain frontal chest radiographs, resulting in an area under the curve of between 0.861 and 0.964 for different chest radiograph features.[14] A different study used transfer learning from a pre-trained ImageNet CNN for the automated categorisation of osteoarthritis in knee radiographs.[15]

Plain radiographs are the most common radiological test with over 22 million studies performed in the UK in 2016.[3] A high proportion of these are plain extremity radiographs in the context of trauma. There is a strong case for developing automated strategies to improve efficiency and workflow management in this area considering the backlog in unreported studies and the fact that over £88 million was spent on outsourcing radiology reports in 2014–2015.[1] It is surprising therefore that, to the author's knowledge, there are currently no studies in the literature that have successfully applied transfer learning from pre-trained CNNs to the problem of fracture detection on plain radiographs. This proof of concept study aims to establish to what extent this is possible.

Materials and methods

This National Health Service (NHS)-based study was granted approval by the Health Research Authority in England. This study was retrospective and used only anonymised data, and therefore, ethics approval was not required.

Imaging study selection

Anonymised lateral wrist radiographs were obtained from the Royal Devon & Exeter Hospital for studies performed between January 2015 and January 2016. Radiographs were excluded if there was a plaster cast in place, if the growth plates of the wrist had not yet fused, or if the study demonstrated any fracture other than a fracture of the distal radius or ulna. The images were classified as either "fracture" or "no fracture" on the basis of the radiological report. This classification was checked and verified by a radiology registrar competent in the reporting of plain radiographs and with 3 years radiology experience. Images were also excluded if the single lateral projection was inconclusive for the presence or absence of fracture.

Image pre-processing

This resulted in a preliminary data set of 695 wrist radiographs demonstrating a fracture and 694 wrist radiographs demonstrating no fracture. The images were converted to JPEG by a trained radiologist ensuring the most appropriate windowing was selected. Images were then pre-processed by removing additional annotations such as the commonly used "Red Spot" annotation applied by radiographers at the time of image acquisition in order to reduce the possibility of over-fitting. Over-fitting describes the process by which the learning algorithm learns features that are not truly representative of the image category in an attempt to reduce the error in the learning process.

A data augmentation technique was used to amplify the data. This involved making a number of non-exact copies, or transformations, of each image. This served to provide the CNN with more training examples by incorporating the salient features in multiple orientations. The aim was to better reflect the real-world population of wrist radiographs,